Cornell University

Exploring the Limits of Semantic Image Compression at Micro-Bits Per Pixel

Jordan Dotzel[1], Bahaa Kotb[1][1], James Dotzel[2], Mohamed Abdelfattah[1], Zhiru Zhang[1]
[1]Cornell University, [2]Penn State University

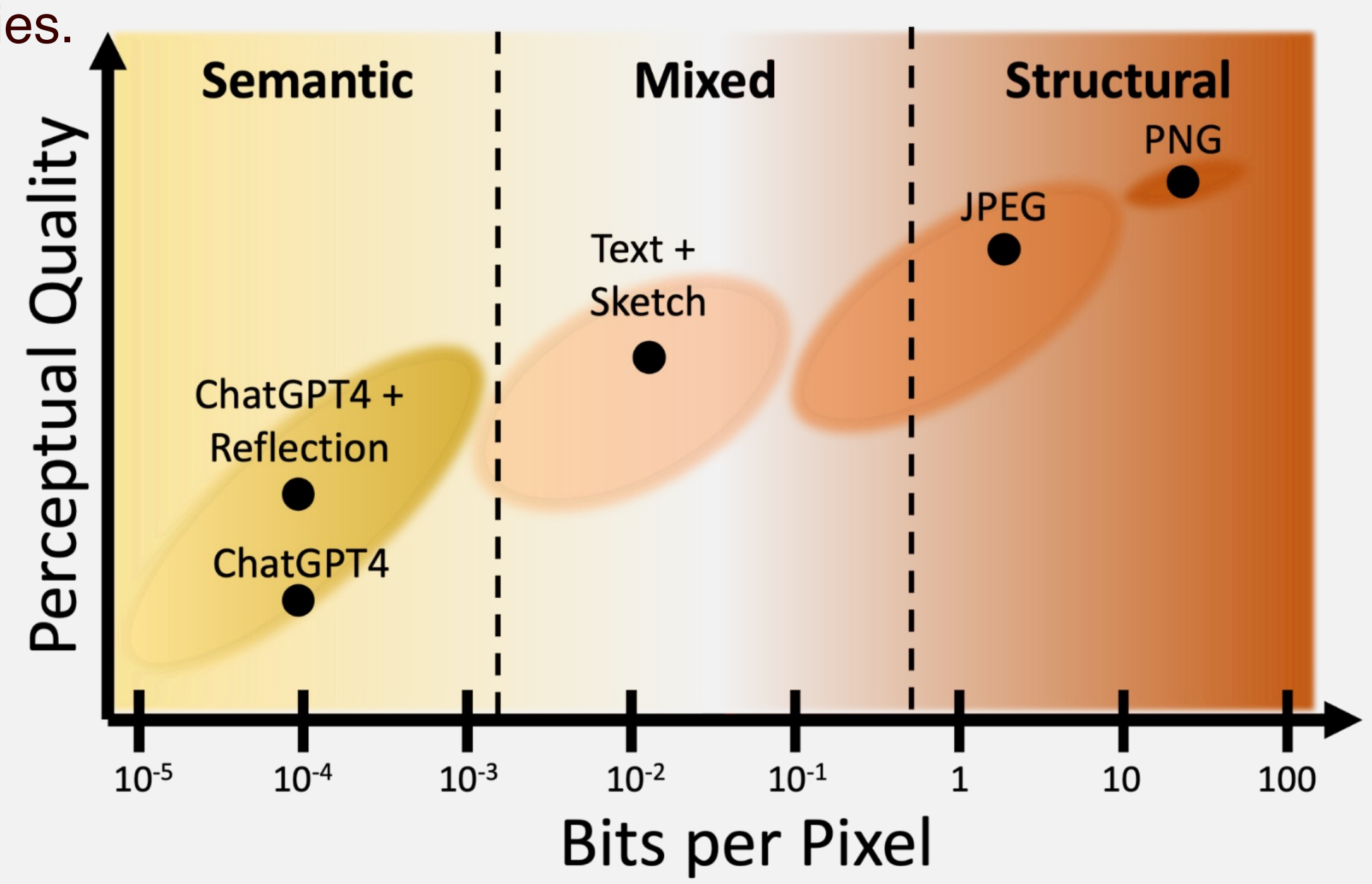
Equal Contribution

Tiny Paper at ICLR 2024

Background

Modern image compression types:

- **Lossless**: Retains all information.
- **Lossy**: Loses some information, transferring data at **lower bitrates** to remove imperceptible details. For instance, **JPEG** retains only key human-perceptible frequencies.



Classification into **three major regions**:

- **Structural region**: Keeps original pixel structure, with JPEG operating between **10^{-1} to 10 bpp**
- **Semantic region:** Focuses on essential human-centric information, achieving compression down to micro-bits per pixel (µbpp).
- **Mixed region**: Merges semantic and structural details, like **Text + Sketch**, which operates at **10^{-3} to 10^{-2} bpp.**

Compression Methodology

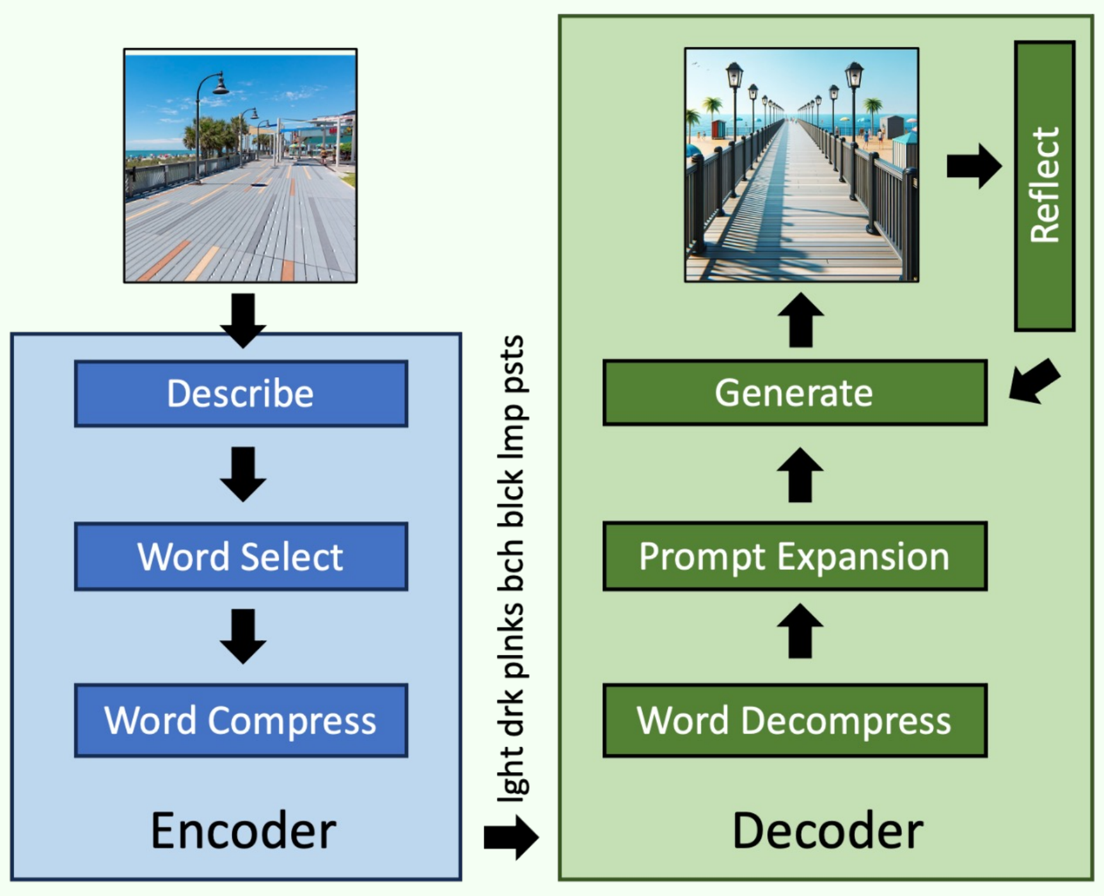
Base Models: Uses **GPT-4V** as encoder and **DALL-E3** as decoder for advanced language compression and image generation.

Encoding Process:

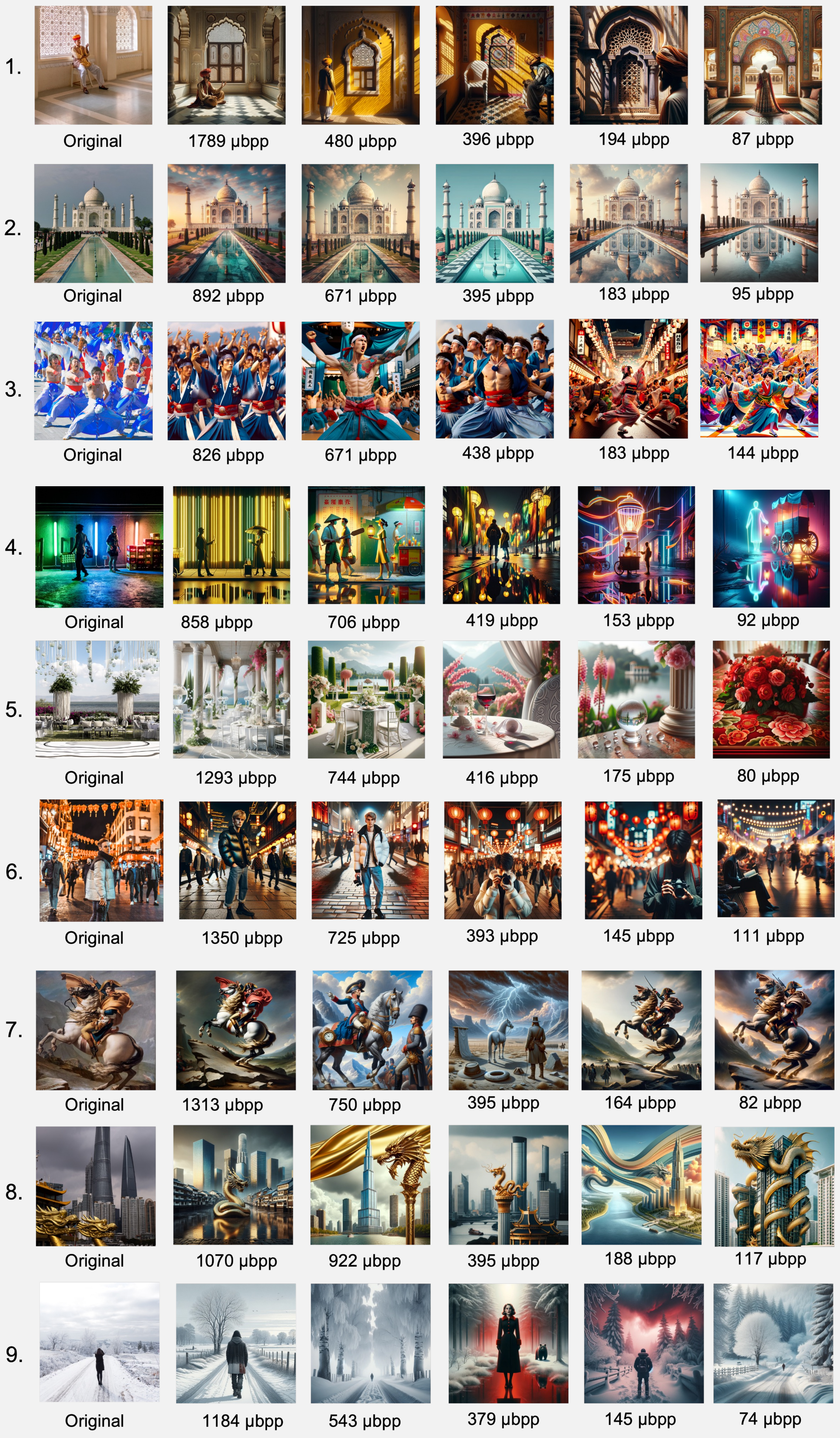
- Analyzes and extracts key details from images, selecting the most important words and compressing them into a **four-bit representation** using only fifteen common consonants.
- **GPT-4V** optimizes text by substituting synonyms and minimizing characters.

Decoding and Image Generation:

- Compressed text is expanded into a natural prompt for **DALL-E3** to generate an image.
- Utilizes an **image reflection technique** for iterative enhancement if the context exceeds **500 µbpp**, typically requiring one or two iterations to refine the image.



Compression Examples



Compression examples demonstrate a **progressive loss of contextual details** like room color and figure position, yet landmarks like the Taj Mahal retain **significant information at extremely low bitrates**, revealing a focus on arbitrary nuanced data like jacket color heavy compression.

Reflection Examples

Word Selection

Indoor ceremony, priest, white robes, golden trim, older, gray hair, light blue surgical mask, serious, offering white communion wafer, left hand, golden chalice, right hand, wafers, white stole, golden patterns, person receiving communion, dark suit, back view, blurred background, bright space, soft diffused light, gentle shadows, muted colors, whites, beiges, white, blue, dark tones contrast.

ndr crmn prst wht rb gldn trm ldr gr hr bl msk srs ffrng wht cmmnn wfr lft hnd gldn chlc rght hnd wfr stl gldn pttrn prsn rcvng cmmnn drk st blck bck vw brght spc sft dffsd lght gntl shdw mtd clrs whts bgs wht bl drk tn cntrst

862 µbpp

Generation

An elder ceremony priest with a robe, golden trimmed, leader, gray hair, blue mask, serious offering, what communion wafer, left hand, golden chalice right hand, wafer still, golden pattern, person receiving communion, dark suit, black back view, bright space, soft diffused light, gentle shadow, muted colors, whites beiges, wheat blue, dark tan, contrast.

Reflection

The most important difference between the two descriptions is the mention of a "blue mask" in the original description which is not mentioned in the first description provided. The suggested change to make to the first description would be to add the detail of the priest wearing a blue mask. Here is how you could integrate it:
"In the foreground, there's a man with gray hair and a neatly trimmed beard, wearing a white robe with golden embroidery and a blue mask, suggesting a priestly garment engaged in a ceremony."

Word Selection

High-angle night urban intersection, zebra crossing, white-dark gray stripes, asphalt contrast. Vehicles: orange-black taxi motion blur, white black sedans stopped, silver sedan, large black van, luxury car moving away. Background: golden-lit high-end store, large glass windows. Streetlamp, white light, center. Perspective: depth, crossing lines to vanishing point. Muted palette, yellowish artificial lighting. No people, city life stillness

nght rbn ntrstn zbr crssng wht drk gr strps sphlt cntrst vhcls rng blck tx mtn blr wht blck sndn stpd slvr sndn lrg blck vn lxry cr mvng wy bckgrnd gldn lt hgh nd str lrgh glss wndw strtlmp wht lght cntr prspctv dpth crssng lns vnshng pnt mtd plt yllwsh rtfl lghtng n ppl cty lf stllnss

1091 µbpp

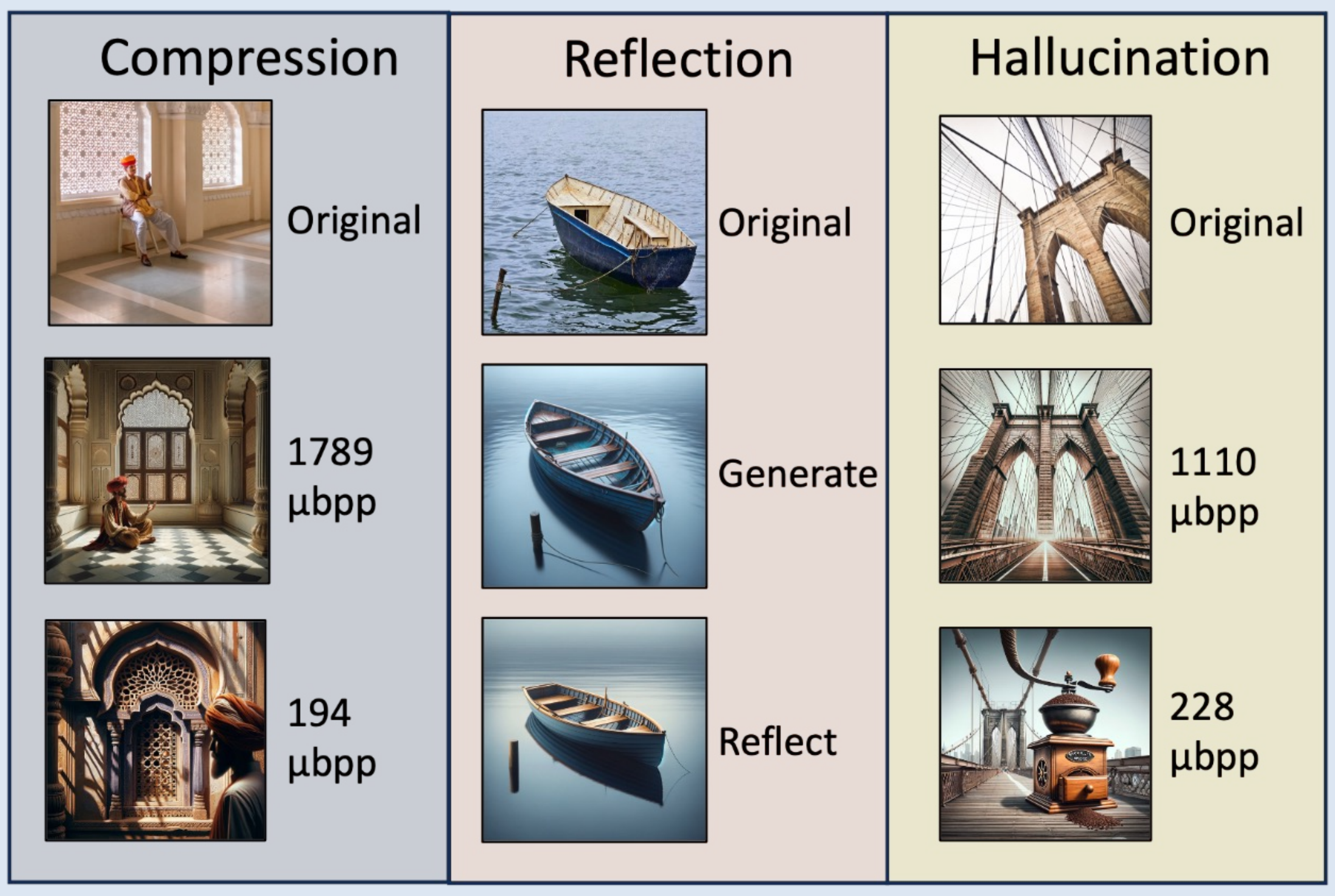
Generation

A night urban intersection with zebra crossing with dark grey stripes on asphalt. In contrast, vehicles ranging from a black taxi, a mountain bike, a white and black sedan, a stopped silver sedan, a large black van, and a luxury car moving away in the background. Golden light is high and scattered, reflecting off large glass windows and streetlamps with white light. Converging perspective and depth with crossing lines and a vanishing point. Muted palette with yellowish artificial lighting and no people, capturing city life's stillness.

Reflection

Reflection noticed the white light in the original description and adjusted.

Experimental Results



ChatGPT4 is a proof-of-concept for useful semantic compression, outperforming DALL-E3's generative abilities. The **practical limit of 100 µbpp was identified**, demonstrating that semantic compression can achieve significant size reduction with manageable loss in quality. The **reflection technique was crucial** for improving fidelity in the reconstructed images.